基于 Lévy 分布的柔软自适应演化采样算法 *

张海鹏，张扬帆，孙　俊

(江南大学 物联网工程学院, 江苏 无锡 214122)

摘　要：已经有相关工作将演化思想引入采样算法中，并结合 Lévy 分布提出了自适应的采样算法。针对 Lévy 分布的参数设置和“厚尾”特性的关系进行了研究，改进了基于 Lévy 分布的演化采样算法，通过设置该分布的参数 α 值为 1.0，1.3，1.7，2.0，分别对应四种转移概率分布，从而增加了生成的候选样本的多样性。理论分析和实验表明，改进算法在收敛速率和精度上优于基于高斯分布、柯西分布，对称指数分布的演化采样算法和其他自适应的演化采样算法。

关键词：演化采样；Lévy 分布；柔软自适应

中图分类号：TP301.6　　doi: 10.3969/j.issn.1001-3695.2017.12.0850

Evolutionary sampling approach with soft adaptive Lévy probability distribution

Zhang Haipeng, Zhang Yangfan, Sun Jun

(*School of Internet of Things Engineering*, *Jiangnan University*, *Wuxi Jiangsu* 214122, *China*)

Abstract: Some research has introduced evolution idea into sampling algorithms, and related algorithms combined with adaptive Lévy distribution are proposed. This paper improves the evolutionary sampling algorithm based on Lévy distribution. By setting the parameter α of this distribution to 1.0, 1.3, 1.7, 2.0, corresponding to the four transition probability distributions, the diversity of the generated candidate samples is increased. Theoretical analysis and experimental results show that the proposed algorithm is superior to the evolutionary sampling algorithm based on Gaussian distribution, Cauchy distribution, symmetrical exponential distribution and other adaptive evolutionary sampling algorithms in terms of convergence rate and accuracy.

Key words: evolutionary sampling; Lévy distribution; adaptive

0　引言

采样是机器学习等相关领域中的重要问题，对于某些概率模型，在精确推断不可行的情况下，采用基于数值采样的近似推断方法可以有效地解决问题。虽然采样并不是一个崭新的研究问题，但是在大数据的背景下，仍然需要进一步研究该问题来满足各种应用场景的需求[1]。

在给定概率分布的前提下，寻找一种有效的采样策略获取符合该概率分布的样本问题已经在学术界得到长期而广泛的研究。线性同余法[2]可以产生满足均匀分布的样本，而反变化采样法[3]基于均匀分布可以产生满足其他分布的样本，但由于线性同余法产生的下一个随机数完全依赖于当前随机数的大小，当随机数序列足够大的时候，可能出现重复子序列。拒绝采样方法利用可直接采样的概率分布通过拒绝某些样本实现对复杂目标概率分布的采样，但是由于寻找合适的可直接采样的概率分布和分布常数比较困难，可能导致该采样方法拒绝率很高。重要性采样是另外一种利用可直接采样的概率分布进行复杂分布采样的方法。马尔可夫链蒙特卡洛方法（Markov chain Monte Carlo, MCMC)[4]建立了生物进化和采样方法之间的联系，通过转移概率对样本进行更新，马尔可夫链收敛后得到的样本就符合目标概率分布。Metropolis-hasting 算法(MH)[5]和 Gibbs 算法[6]是基于 MCMC 的两个采样算法，MH 算法改进了马尔可夫链状态转移过程中接受率可能偏小的问题，而 Gibbs 算法则是针对多维采样而提出。有人将基于 MCMC 的采样算法应用到 K-means 初始种子点的选择问题上[7; 8]，取得了良好的效果。近期，谢，孙等人[9]提出的演化采样算法 (Evolutionary Sampling, ES) 结合拒绝采样和进化优化并继承了二者的优点。除此之外，针对不同的应用场景，还有很多采样算法被提出[10-12]。

Lévy 分布在进化优化领域得到了充分的研究[13; 14]，胡[15]将 Lévy 分布作为可直接采样的概率分布引入演化采样算法中，取得了比采样分布为高斯分布，柯西分布，对称指数分布更高的精度和更快的收敛速率，但是胡只考虑了 Lévy 分布参数 α 在特殊情形下的应用，没有考虑一般的情况。本研究通过综合考虑特殊和一般情况，提出基于 Lévy 分布的柔软自适应演化采样算法，理论和实验均证明了该算法在精度和收敛速率上优于胡的工作。

收稿日期：2017-12-12；**修回日期：**2018-03-02　　**基金项目：**国家自然科学基金资助项目（61672263）

作者简介：张海鹏（1991-），男，河南巩义人，硕士研究生，主要研究方向为数据挖掘与机器学习（1332062790@qq.com）；张扬帆（1994-），女，硕士研究生，主要研究方向为机器学习；孙俊（1971-），男，教授，博士，主要研究方向为人工智能、机器学习、计算智能和高性能计算.

1 演化采样算法和 Lévy 分布

1.1 采样理论

机器学习中的概率模型要解决的基本问题之一是关于一个概率分布 $p(z)$ 寻找某个函数 $f(z)$ 的期望，在连续变量的情形下，也就是计算期望

$$\mathbb{E}[f]=\int f(z)p(z)dz \tag{1}$$

在离散变量的情形下，积分被替换为求和。在使用解析方法计算期望比较复杂的情况下，可以使用采样的方法来计算。采样方法的一般思想是从概率分布 $p(z)$ 中独立抽取一组变量 $z(l)$，其中 $l=1,2,\ldots,L$，这使得期望可以通过有限和的方式计算，即

$$\hat{f}=\frac{1}{L}\sum_{l=1}^{L}f(z^{(l)}) \tag{2}$$

只要满足样本 $z(l)$ 是从概率分布 $p(z)$ 中抽取，那么 $\mathbb{E}[\hat{f}]=\mathbb{E}[f]$，但如果要满足 $\hat{f}=\mathbb{E}[f]$，需要准确地获取满足 $p(z)$ 概率分布的样本集 $z(l)$，因此如何准确地获取满足某个给定分布的样本是一个重要问题。

1.2 演化采样方法

演化采样算法是谢，孙等人[9]受进化计算启发提出的一种新的采样方法，适用于解决概率框架下的机器学习问题。支持样本模型(support sample model, SSM)是 ES 中的一种计算模型，目的是近似样本数据的概率密度分布，因此演化采样方法的实现需要寻找 SSM 的最优参数。SSM 利用一组高斯密度函数的加权组合来表示样本集的概率密度分布，定义为

$$\hat{p}(x)=\sum_{i}\alpha_i K_h(x-\mathbf{x}_i) \tag{3}$$

$\mathbf{x}_i$ 是样本数据,其中 $i=1,2,\ldots,N$，N 是样本的数量，α_i 是权值系数,值为 $1/N$,高斯核函数被定义为

$$K_h(x;h)=\frac{1}{h\sqrt{2\pi}}exp\left\{-\frac{\|x\|}{2h^2}\right\} \tag{4}$$

其中 $\|\cdot\|$ 是二范数，h 是带宽参数。因此，ES 的目标在概率框架下可以表示为

$$X_s^*=arg_{x_s}min\left\{\|\hat{p}(\cdot)-\pi(\cdot)\|_{TV}\right\} \tag{5}$$

X_s^* 是所有可能支持样本的最优组合，$\hat{p}(\cdot)$ 和 $\pi(\cdot)$ 分别表示 SSM 和目标模型的概率分布，$\|\cdot\|_{TV}$ 表示全变量距离。上式意味着最优的支持样本集可以在演化过程中通过使 SSM 和目标模型之间的距离最小化得到。ES 算法根据拒绝采样策略的核心思想，为每个支持样本设计相似的采样程序。对每个采样链，让生成的候选样本以一定的概率被接收成为新的样本。令 $A^k(x,y)$ 表示在第 k 个拒绝采样步骤中，从样本 x 生成新的候选样本 y 的概率，$\alpha^k(x,y)$ 表示接收数据 y 作为新样本的概率，则如果当前支持样本数据是 x，在第 k 个拒绝采样步骤之后，接收样本数据 y 作为新样本的概率是 $A^k(x,y)\times\alpha^k(x,y)$。等式(5)的求解依据下述表达式

$$A^k(x,y)=\begin{cases}q^k(x,y)\times\alpha_{MH}(x,y), x\neq y\\ \int q^k(x,t)\times[1-\alpha_{MH}(x,t)]dt, x=y\end{cases} \tag{6}$$

$$\alpha^k(x,y)=max\left\{1-\frac{\pi(x)\times p^k(y)}{\pi(y)\times p^k(x)},0\right\} \tag{7}$$

其中：$q^k(x,y)$ 是从 x 生成新的候选样本 y 的概率，满足 $q^k(x,y)=q^k(y,x)$，ES 中 $q^k(x,y)$ 的定义如下：

$$q^k(x,y;\beta)\propto e^{-\frac{|x-y|}{\beta}} \tag{8}$$

这是一个对称指数分布，β 是搜索规模因子，满足 $\beta=\beta_1\frac{1}{N_s}\sum_{x_i\in X_s}\|x_i-\bar{x}\|$，$N_s$ 表示支持样本集中的样本数量，β_1 的值在演化过程中从 20 到 1 线性递减[16; 17]。$\alpha_{MH}(x,y)=min\left\{1,\frac{\pi(y)}{\pi(x)}\right\}$ 满足细致平稳条件，即 $\pi(x)A^k(x,y)=\pi(y)A^k(y,x)$，不同于传统的拒绝采样策略，$\alpha^k(x,y)$ 是由谢，孙等人首先提出并应用在演化采样算法中的拒绝选择法则，是 ES 算法的第二次拒绝选择操作，能够确保 $p^k(x)$ 对 $\pi(x)$ 的近似效果，而 $p^k(x)$ 正是上文所定义的支持样本模型。

1.3 Lévy 分布

Lévy 分布是由 P.Lévy 在 1937 年提出的一种具有无限二阶矩的概率分布，其形式如下：

$$L_{\alpha,\gamma}(y)=\frac{1}{\pi}\int_0^\infty e^{-\gamma q^\alpha}cos(qy)dq, y\in R \tag{9}$$

当 $y=0$ 时，该分布满足对称性，$\gamma(>0)$ 是缩放因子，α 决定该分布的形状，满足 $0<\alpha<2$。

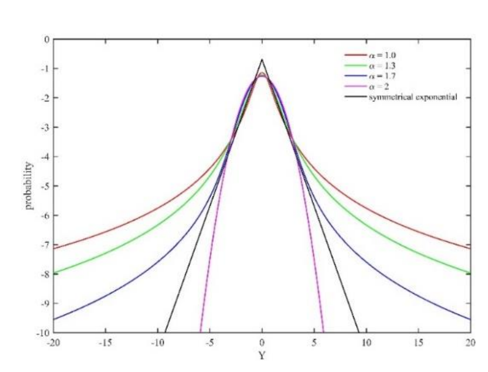


图 1 α 分别为 1.0,1.3,1.7,2.0 时 Lévy 分布和对称指数分布的比较，纵坐标做了 log 变换

式(9)在一般情况下的积分结果是未知的，但是在两个特殊情况下，可以知道对应的积分结果。$\alpha=1$ 时，对应柯西分布，$\alpha\to 2$ 时，该分布变为高斯分布。α 控制分布的形状，特别是分布的“厚尾”区域，α 越小，“厚尾”越长。图 1 显示了不同 α 值对应的 Lévy 分布和对称指数分布的比较结果。

1.4 转移概率分布比较

转移概率的搜索空间是影响生成样本质量的关键因素，通

过对比高斯分布，对称指数分布，Lévy 分布，从均方位移的角度来分析三种分布搜索空间的大小。父样本和子样本之间的关系可表示为 $x^t = x^{t-1} + \beta\eta^{t-1}$，其中 x^t 表示第 t 代的样本，η 是给定概率分布产生的随机变量，β 是一个常数。用 $\Delta(t)$ 表示样本 x 在经过 t 代的更新后产生的变化量，则 $\Delta(t) = x^t - x^0 = \beta\sum_{k=0}^{t-1}\eta^k$。对于高斯分布，$\eta^t \sim N(0,1)$，则 $\Delta(t) \sim N(0,\beta^2 t)$，均方位移为 $\sqrt{\langle\Delta(t)^2\rangle} = \beta\sqrt{t} \propto \sqrt{t}$，其中 $\langle\cdots\rangle$ 表示期望。类似地，对于对称指数分布可以得到 $\sqrt{\langle\Delta(t)^2\rangle} = \beta\sqrt{t^2 - t} \propto t$，不失一般性，令 $\gamma = 1$，$L_{\alpha,\gamma} = L_\alpha$，则 Lévy 分布可以被近似表示[18]为

$$L_\alpha(y) \propto y^{-(\alpha+1)}, |y| \gg 1 \tag{10}$$

可以得到

$$\langle y^2\rangle = \int_{-\infty}^{\infty} \frac{y^2}{y^{\alpha+1}} dx = \infty, \alpha \in (0,2) \tag{11}$$

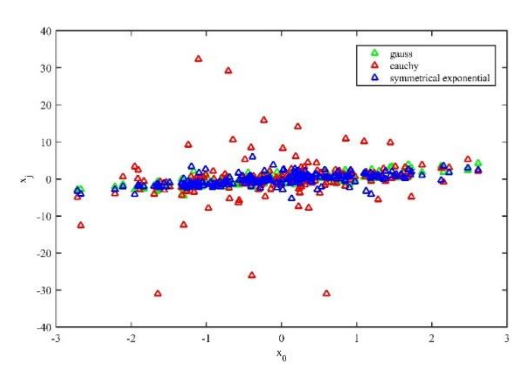


图 2　基于高斯分布，柯西分布和对称指数分布从父代产生的子代的分布情况对比

柯西分布是 Lévy 分布在 $\alpha = 1$ 时的对应分布。从上述分析可以看出，对一个特定样本产生相同的变化量，高斯分布需要时间最长，对称指数分布次之，而 Lévy 分布能够在短时间产生大的变化量。令 $t = 1$，$\beta = 1$，由 $x^t = x^{t-1} + \beta\eta^{t-1}$ 得 $x^{i1} = x^{i0} + \eta^{i0}, i = 0,1,2$，$i$ 的值分别代表高斯分布，对称指数分布和柯西分布，η^{i0} 是分别由三者生成的随机变量。从图 2 可以看出，柯西分布产生的变量比高斯分布和对称指数分布能够覆盖一个更宽的范围。

2　基于 Lévy 分布的柔软自适应演化采样

从上述分析可以看出，高斯分布可以产生小范围的变化量，有助于通过局部搜索提高精度，柯西分布可以产生大范围的变化量，有助于通过全局搜索跳出局部最优。胡[15]的工作受 X Yao[18]等人的启发，通过设定 Lévy 分布中的 α 值分别为 1.0 和 2.0，为父代样本产生两个候选样本，选择最好的样本作为下一代样本。Lévy 分布在 $\alpha = 1.0$ 和 $\alpha = 2.0$ 时的分布，实际上分别对应了柯西分布和高斯分布两个极端分布。在上文中提到，参数 α 控制 Lévy 分布的形状，特别是分布的“厚尾”区域，参数 α 越小，“厚尾”区域越长。较小的 α 可以产生更大的搜索空间，较大的 α 可以实现更好的局部搜索。通过添加两个 α 值分别为 1.3 和 1.7 的一般 Lévy 分布，增加另外两个候选样本，使得候选样本的选择不会局限于柯西分布和高斯分布，从而增加了选择的多样性。这是一种更加柔软的自适应选择方法，算法具体过程如下：

a)设置样本个数 N_s 和最大迭代次数 K_{max}，初始样本集为 $X_s^0 = \{X_1^0, X_2^0, \cdots, X_{N_s}^0\}$，初始化迭代次数 $k = 0$；

b)对 X_s^k 中的每个样本 X_i^k，执行步骤 $3\sim6$；

c)利用 Lévy 分布作为概率转移函数 $q^k(x_i^k, y)$，对应不同 α 值生成四个候选样本 y_{ij}^C，其中 $j = 1,2,3,4$，分别对应 $\alpha = 1.0, 1.3, 1.7, 2.0$，分别执行步骤 d)~f)；

d)生成 [0,1] 内的一个均匀随机数 R，如果 $R < min\left\{1, \frac{\pi(y_i^C)}{\pi(x_i^k)}\right\}$，接受 y_i^C 后执行下一步骤；否则拒绝 y_i^C，执行步骤 f)；

e)根据等式 (7) 计算 $\alpha^k(x_i^k, y_i^C)$，生成 [0,1] 内的另一个均匀随机数 R'，如果 $R' < \alpha^k(x_i^k, y_i^C)$，接受 y_i^C，否则拒绝；

f)如果 y_i^C 被接受，则令 $x_i^{k+1} = y_i^C$；否则令 $x_i^{k+1} = x_i^k$；

g)令 $k = k+1$，分别更新样本集为 $X_{sj}^k = \{X_{j1}^k, X_{j2}^k, \cdots, X_{jN_s}^k\}$，计算对应标准偏差 J_j^k，其中 $j = 1,2,3,4$，分别对应 $\alpha = 1.0, 1.3, 1.7, 2.0$，选择最小的 J_j^k 对应的样本集 X_{sj}^k 为选择的样本集；

h)如果连续两个样本集之间差异很小或者达到最大迭代次数 K_{max}，结束程序；否则转步骤 b)；

i)输出最终样本集 X_s^k，结束程序。

为了对算法的时空复杂度进行分析，对算法的数学符号进行了简化。假设样本个数为 N，最大迭代次数为 K，标准偏差为 J。对每个样本进行两次拒绝选择操作需要常数时间，每个迭代过程需要进行 4N 次拒绝选择操作，同时可以在常数时间选择最小偏差 J 对应的样本集，故每个迭代过程的时间复杂度为 O(N)，总的时间复杂度为 O(K*N)，算法的空间占用与样本个数 N 呈线性增长，故空间复杂度为 O(N)。考虑到算法的并行性，通过并行编程加速采样也是下一步要进行的工作。

3　实验和结果分析

3.1　实验设置

“拇指法则”是由 Deheuvels[19]提出,Silverman[20]推广的可用于生成高斯核函数中近似带宽的有效方法。Silverman 拇指法则表达式如下：

$$h = \left(\frac{4}{3N}\right)^{1/5} \hat{\sigma} \tag{12}$$

其中：N 是样本数量，$\hat{\sigma}$ 是样本集的平均绝对值偏差。为了和胡[15]的工作进行对比，在实验中使用相同的方法进行带宽参数的选择。为了验证所提算法的改进效果，做了 5 个 ES 算法的

变种在三类测试概率函数上的实验，实验中用到的 ES 算法变种包括：采样分布函数为柯西分布的演化采样算法(ES with Cauchy probability distribution, ES-CPD)；采样分布函数为对称指数分布的演化采样算法(ES with symmetrical exponential probability distribution, ES-SEPD)；采样分布函数为高斯分布的演化采样算法(ES with Gaussian probability distribution, ES-GPD)；采样分布为两点自适应 Lévy 分布的演化采样算法(ES with two points adaptive Lévy probability distribution, ES-HARD-ALPD)；采样分布为四点自适应 Lévy 分布的演化采样算法 (ES with two points adaptive Lévy probability distribution, ES-SOFT-ALPD)；选择的三类测试函数如表 1 所列，其中是 $\pi_1(x)$ 单峰函数，$\pi_2(x)$ 是具有轻微局部最优的函数，$\pi_3(x)$ 是具有明显局部最优的函数。

为了比较 5 个 ES 算法，对每个测试函数执行 30 次采样，初始 N_S 个样本通过标准正态分布产生。标准偏差 J 的定义为 $J=\sqrt{\sum_{i=1}^{N}(p(x_i)-\pi(x_i))^2}$，其中 $p(x)$ 表示标准化输出，$\pi(x)$ 表示目标分布，J 用来衡量所得到的样本集和目标概率分布样本集的接近程度，该值越小，表示精度越高。定义标准化因子 $\gamma=\frac{1}{N_s}\sum_{x_i\in X_s}\frac{\pi(x_i)}{p(x_i)}$，该值用来评价算法的收敛性，当 γ 收敛到一个稳定的值时，ES 算法即收敛得到一个最优的稳定解。此外，在计算 J 和 γ 的时候，从表 1 所示变量的取值范围中均匀选取 1000 个样本点。

表 1　测试函数集

$\pi(x)$	$Range(x)$	D
$\pi_1(x)=\frac{1}{8}\frac{1}{\sqrt{2\pi}}\sum_{i=0}^{7}(\frac{1}{\sqrt{h_i}}e^{-(x+3-3h_i)^2/2/h_i}), h_i=(\frac{2}{3})^i$	$(-5,20)$	1
$\pi_2(x)=(1+x+2x^2)e^{-x^2}$	$(-5,5)$	1
$\pi_3(x)=0.7\pi_1(x)+0.3\frac{1}{3\sqrt{2\pi}}e^{-(x-16)^2/2/9}$	$(-5,20)$	1

3.2　实验结果

为了比较改进算法和其他算法之间的精度差异，基于最后一次迭代执行完成后的 J 值做了 t 检验，显著性水平设为 0.05，自由度为 58。检验结果如表 2 所示，其中“+”表示存在显著性差异，“-”表示不存在显著性差异。最后一次迭代执行完成后五个算法各自得到的 J 和 γ 的平均值分别如表 2 和 3 所示。从表 2 可以看出，所改进算法和其他对比算法在精度上存在显著性差异，并且在三类测试函数上的 J 值小于其他算法，说明该算法精度高于其他算法。从表 3 可以看出所提算法在三类测试函数上最后一次迭代执行完成后 γ 值小于 1，说明该算法收敛。

整个进化过程的具体实验结果如图 3 所示，对于每一类函数，依次显示 J，γ，进化过程中接受样本数量和 ES-SOFT-ALPD 算法中不同 α 值对应的接受样本数量。对于 $\pi_1(x)$，从图 3 可以看出，ES-SOFT-ALPD 算法和 ES-HARD-ALPD 算法在精度上接近并高于其他三个算法。ES-GPD 算法利用高斯函数作为概率转移函数，具有最小的搜索范围，整体上收敛速度较慢，从演化的初期开始接受大量的候选样本，但是经过很长的时间才收敛到样本的最优状态。而 ES-CPD 算法，由于 $\pi_1(x)$ 是单峰函数，从收敛曲线图可以看出，在短时间内可以快速收敛，但是由于柯西分布的搜索步长较长，候选样本成功接受的数量并不多，使得在演化过程中接近最优状态时，可能跳过最优状态，在标准偏差图中可以看到这一现象。ES-SEPD 算法通过对称指数函数的参数在一定范围内进行线性调节，使搜索范围从大到小逐渐变化，保证演化算法实现从全局搜索到局部搜索，在收敛和精度上均优于 ES-GPD 算法。从整体来看，ES-SOFT-ALPD 表现的最好。在演化初期，利用柯西分布进行大范围搜索，接收候选样本数量较少，中期利用 Lévy 分布逐步减小搜索范围，随着演化的进行，ES-SOFT-ALPD 使用高斯分布在更小的范围进行搜索，接受尽可能多的更加优秀的候选样本，提高算法精度，所以 ES-SOFT-ALPD 在保证收敛的同时，能够得到比其他算法更高的精度。

ES-HARD-ALPD 在演化过程中使用 Lévy 分布的两个极限分布，使得搜索范围在最大和最小之间切换，忽略了进化过程中间的搜索范围，从而影响到优秀候选样本的接受，故在精度上比 ES-SOFT-ALPD 略差。为了进一步研究不同 α 值在 ES-SOFT-ALPD 算法中的作用，从不同 α 值对应的样本接受数量图中可以看出，对于单峰函数 $\pi_1(x)$，不同 α 值对应的分布均在整个演化采样过程中被用到。

表 2　最后一次迭代完成后 J 的值

函数	ES-GPD	ES-SEPD	ES-CPD	ES-HARD-ALPD	ES-SOFT-ALPD
$\pi_1(x)$	16.5829(+)	15.1066(+)	14.4624(+)	8.6283(+)	**8.2348**
$\pi_2(x)$	5.7115(+)	6.2795(+)	7.1493(+)	4.0108(+)	**3.0300**
$\pi_3(x)$	14.7393(+)	2.1490(+)	2.0347(+)	1.5830(+)	**1.2783**

表 3　最后一次迭代完成后 γ 的值

函数	ES-GPD	ES-SEPD	ES-CPD	ES-HARD-ALPD	ES-SOFT-ALPD
$\pi_1(x)$	1.2115	1.1070	1.0341	0.7978	0.8039
$\pi_2(x)$	0.5296	0.4459	0.4058	0.4982	0.5604
$\pi_3(x)$	0.0646	0.4599	0.5078	0.5359	0.5769

$\pi_2(x)$ 是具有轻微局部最优的函数，从图 3 可以看出，ES-CPD 算法和 ES-SEPD 算法产生的效果类似，但 ES-SOFT-ALPD 算法和 ES-HARD-ALPD 算法在精度上更高，同时 ES-SOFT-ALPD 算法进一步提高了 ES-HARD-ALPD 算法的精度。由于 $\pi_2(x)$ 具有轻微的局部最优特性，所以较大范围的搜索有助于跳出局部最优，产生好的近似效果。ES-SEPD 对于不同的问题需要对参数进行线性调整，其搜索范围可能会大于柯西分布，故从收敛曲线图中可以看到 ES-SEPD 具有最快的收敛速率，但是由于演化初期成功接受的样本数量并不多，故在精度

上略差。样本接受数量图表明，ES-SOFT-ALPD 算法虽然在演化初期成功接受的样本数量略少于 ES-HARD-ALPD 算法，但是在中后期比 ES-HARD-ALPD 多，因此在收敛速度相似的前提下，精度上高于 ES-HARD-ALPD 算法。同时从不同 α 值对应的样本接受数量图中看出，在演化过程初期，主要是柯西分布发挥在候选样本选择时的作用，这有助于快速收敛到最优样本集，在演化的中后期，主要利用高斯分布逐渐提高样本集的精度，这个过程中 $\alpha=1.3$ 和 $\alpha=1.7$ 对应的分布同样在样本选择时发挥了重要作用。演化初始阶段，初始样本集和最优样本集的全变量距离较远，柯西分布利用一个大范围的搜索，有助于跳出局部最优，随着演化过程的进行，全变量距离也在缩短，此时，通过高斯分布进行小范围搜索，提高样本精度。

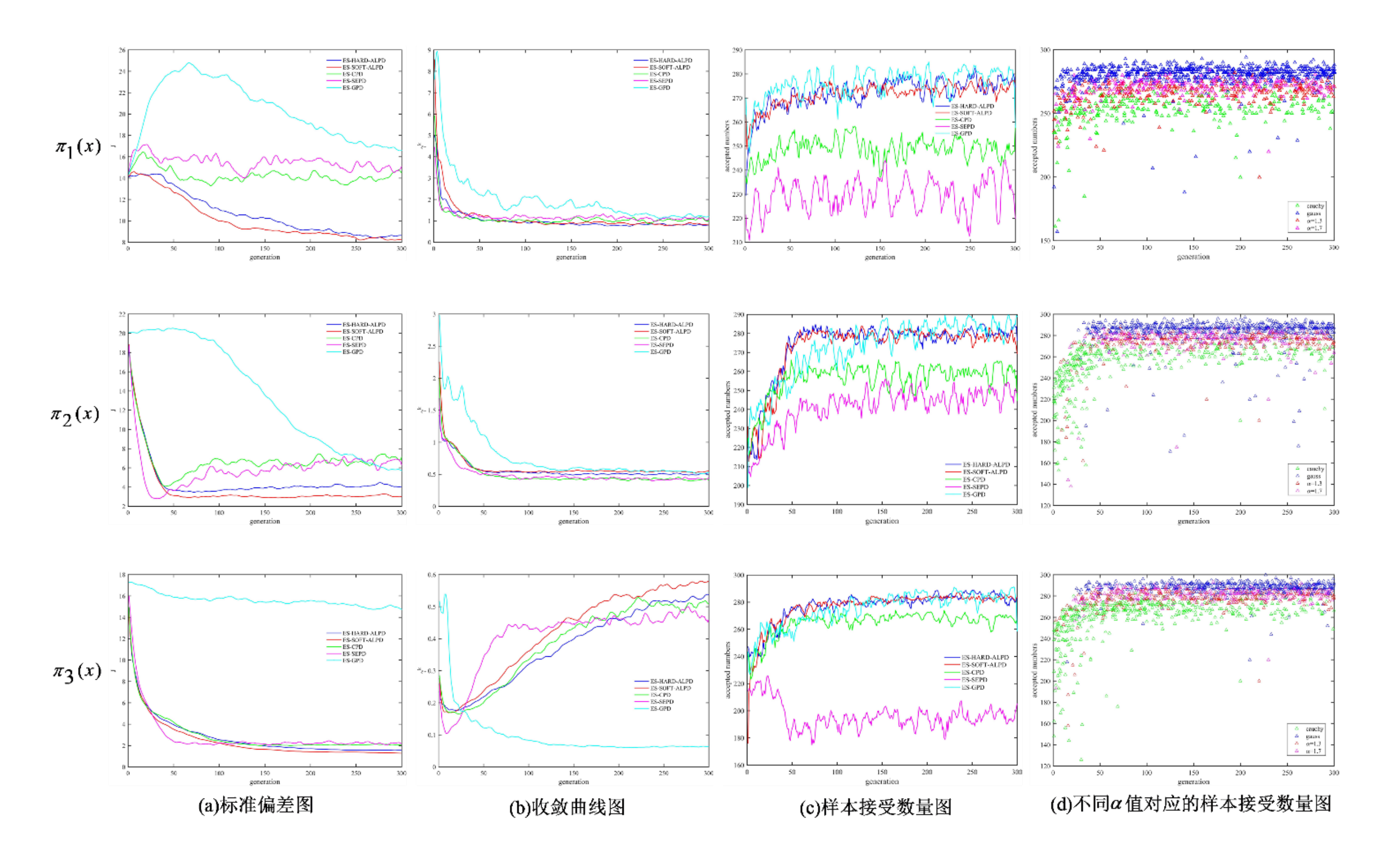


图 3　五种算法在三类测试函数上的实验结果对比图

对于具有明显局部最优的测试函数 $\pi_3(x)$，ES-SOFT-ALPD 也可以产生较好的近似效果，能够产生更大搜索范围的概率转移函数可以使演化采样算法跳出局部最优。从整体上来看，ES-SOFT-ALPD 算法，ES-HARDALPD 算法，ES-CPD 算法和 ES-SEPD 算法在精度上远远高于 ES-GPD 算法，ES-GPD 由于采用高斯分布进行小范围的搜索，容易陷入局部最优，故从样本接受数量图中看出，虽然在整个进化过程中 ES-GPD 成功接受的样本数量并不低于其他算法，但 ES-GPD 具有最差的精度表现。ES-HARD-ALPD 综合利用高斯分布和柯西分布，在精度上优于 ES-SEPD 和 ES-CPD 算法，但是由于更多样的 α 值选择，ES-SOFT-ALPD 进一步提高了 ES-HARDALP 的精度。不同 α 值对应的样本接受数量图说明对于具有明显局部最优的函数 $\pi_3(x)$，四种对应分布在演化过程中发挥的作用和轻微局部最优的函数 $\pi_2(x)$ 类似，再一次验证了 $\alpha=1.3$ 和 $\alpha=1.7$ 对应的分布在候选样本选择过程中的意义。

4　结束语

本文提出了一种更加柔软的自适应演化采样算法，理论分析和实验结果均证明了在精度和收敛速度上优于基于高斯分布，柯西分布，对称指数分布的演化采样算法和其他自适应演化采样算法，但是没有考虑 $0<\alpha<1.0$ 的情况，原因是在此区间 Lévy 分布产生的随机数的范围过大。此外，采用的是离散形式的自适应方案，因此连续形式的自适应方案也是需要进一步研究的问题。最后，研究新型的能够提高演化采样算法的收敛速度和精度的采样分布也是下一步将要进行的重要工作。

参考文献：

[1] Tsai C W, Lai C F, Chao H C, *et al.* Big data analytics: a survey [J]. Journal of Big Data, 2015, 2 (1): 21.

[2] Fontaine C. Linear congruential generator [J]. Encyclopedia of Cryptography & Security, 2005, (3): 480-484.

[3] Olver S, Townsend A. Fast inverse transform sampling in one and two dimensions [J]. Statistics, 2013.

[4] 朱新玲. 马尔可夫链蒙特卡罗方法研究综述 [J]. 统计与决策, 2009, (21): 151-153.

[5] Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm [J].

American Statistician, 1995, 49 (4): 327-335.

[6] Kim Changjin, Nelson C R. State-space models with regime-switching: classical and gibbs sampling approaches with applications by [M]. Cambridge: MIT Press, 1999: 105.

[7] Bachem O, Lucic M, Hassani S H, *et al.* Approximate k-means+in sublinear time [C]// Pro of the 30th AAAI Conference on Artificial Intelligence. 2016: 1459-1467.

[8] Bachem O, Lucic M, Hassani H, *et al.* Fast and provably good seedings for k-means [C]// Advances in Neural Information Processing Systems. 2016: 55-63.

[9] Xie Z, Sun J, Palade V, *et al.* Evolutionary sampling: a novel way of machine learning within a probabilistic framework [J]. Information Sciences, 2015, 299: 262-282.

[10] 李龙洋，董一鸿，严玉良，等. Spark 环境下基于频繁边的大规模单图采样算法 [J]. 计算机研究与发展, 2017, 54 (9): 1966-1978.

[11] Billio M, Casarin R, Osuntuyi A. Efficient Gibbs sampling for Markov switching GARCH models [J]. Computational Statistics & Data Analysis, 2012, 100: 37-57.

[12] 古平，杨炀. 面向不均衡数据集中少数类细分的过采样算法 [J]. 计算机工程, 2017, 43 (2): 241-247.

[13] Yao X, Liu Y, Lin G. Evolutionary programming made faster [J]. IEEE Trans on Evolutionary Computation, 2002, 3 (2): 82-102.

[14] Lee C Y, Yao X. Evolutionary algorithms with adaptive Levy mutations [C]// Proc of Congress on Evolutionary Computation. 2001: 568-575.

[15] Hu C, Sun J, Na T. An evolutionary sampling approach with adaptive Levy probability distribution [C]// Evolutionary Computation. 2016: 1991-1997.

[16] Sun J, Xu W, Feng B. A global search strategy of quantum-behaved particle swarm optimization [C]// Proc of IEEE Conference on Cybernetics and Intelligent Systems. 2004: 111-116.

[17] Wei F. Convergence analysis of quantum-behaved particle swarm optimization algorithm and study on its control parameter [J]. Acta Physica Sinica, 2010, 59 (6): 3686-3694.

[18] Lee C Y, Yao X. Evolutionary programming using mutations based on the Levy probability distribution [J]. IEEE Trans on Evolutionary Computation, 2004, 8 (1): 1-13.

[19] Deheuvels P. Estimation non paramétrique de la densité par histogrammes généralisés [J]. Rev. Statist. Appl, 1977, 25 (3): 5-42.

[20] Silverman B W. Density estimation for statistics and data analysis [M]. [S. l.] : CRC Press, 1986.